

数字资源保存的研究进展、热点与前沿*

□ 胡泽文 / 南京大学信息管理系 南京 210093
/ 中国科学技术信息研究所 北京100038
武夷山 / 中国科学技术信息研究所 北京 100038
孙建军 / 南京大学信息管理学院 南京 210093

摘要: 文章首先辨析了数字资源保存的概念, 介绍了数字资源保存的发展史和国内外研究现状。然后利用CitespaceII可视化分析工具和方法构建领域知识基础(指引文)的时间序列演化图谱和时区演化图谱, 并结合绘制的领域发展趋势图, 分析了数字资源保存的研究进展。同时利用CitespaceII可视化分析工具和方法绘制术语与关键词的混合共现网络和引文共被引聚类知识图谱, 识别出数字资源保存的研究热点^①, 并利用爆发词或突变词探测方法探测数字资源保存的研究前沿。最后针对图书情报与档案管理学科关注最多的研究热点——数字资源长期保存, 从研究主题、最佳实践、保存周期、保存技术和方法、保存策略等几个重要方面解读数字资源长期保存的研究进展。

关键词: 数字资源保存, 研究进展, 研究热点, 研究前沿, 知识图谱, 数字资源长期保存

DOI: 10.3772/j.issn.1673—2286.2013.02.003

1 引言

数字资源保存是图书情报档案学、哲学、法学、经济学和计算机科学交叉的一个研究领域, 是当代情报学发展的重要方向之一, 联合国教科文组织《全民信息计划2008—2013 年战略规划》^②将其列为五大优先发展领域之一。

随着互联网应用的不断深入和信息技术的快速发展, 海量的数字化内容正不断被创造出来, 其中很多内容非常有价值, 是我们文化和知识遗产的一部分。然而互联网应用和信息技术的更新换代周期越来越短, 之前被创造的数字内容会随着一些过时的互联网应用和数字化

技术、软硬件环境变化或载体故障而面临退化、荒废和流失的风险。

因此, 我们需要制定数字保存计划和政策, 根据数字资源的生命周期长短、类型和价值大小等, 开发出一些数字资源保存技术, 对一些有价值的数字资源进行保存, 以备用户将来需要时能够方便获取。

目前国内外已经出现一些数字资源保存方面的综述文章。美国学者Lopez-Fitzsimmons B和英国学者Day M在2003年分别撰文阐述了数字保存和元数据的历史、理论和实践(2002年, 4位学者: Law DG、Sill LA、Gilliland-Swetland A和Chad B也分别发表了4篇相同篇名的文章)^[1,2]; 西班牙学者Termens,

M在2009年对国际数字保存研究和发展进行一个调查研究, 文章概述了机构知识库、政务信息、个人档案、航空航天工业和博物馆档案等领域中的数字保存应用, 重点论述了图书馆和档案馆中的数字保存应用, 最后总结出国际研究的主线包括集成化保存系统构建工具的集成^[3]。武汉大学信息管理学院学者刘家真于2006在《中国图书馆学报》上发表一篇文章对我国数字资源保存的现状进行一个调查研究^[4]。中国科学院国家科学图书馆的吴振新、刘建华、李丹和向菁等学者在2007年和2009年分别对2007年和2008年举办的数字资源长期保存国际会议(iPRES)进行综述, 介绍了数字

* 基金项目: 国家自然科学基金项目(编号: 70973118)和江苏省普通高校研究生科研创新计划项目(编号: CXZZ12_0075)。

① 研究前沿是近几年刚兴起的, 具有前瞻性、先导性和探索性的研究领域或主题, 而研究热点是指一定时期内大家关注比较多的, 稳定集中的研究领域或主题, 两者有交叉。

② http://portal.unesco.org/ci/en/files/26352/12114609343/fifap_draf_strategic_plan.pdf/fifap_draf_strategic_plan.pdf.

资源长期保存的当前进展、最佳实践、保存工具及方法^[5,6]。不过目前笔者还未发现有人利用可视化知识图谱分析方法对数字资源保存相关文献进行全面系统性分析。

2 数字资源保存的概念及不同观点

数字资源是以电子数据的形式将文字、图像、图形、声音、动画等多种形式的信息存储在光磁等载体的数字编码信息,是人们通过计算机可以利用的信息的总和^[7]。数字资源包括数字化资源和原生数字资源:数字化资源是指传统非数字化信息源如印刷型信息源、缩微型信息源和声像型信息源等通过扫描或其他数字化手段转化成相似等价物的数字对象;原生数字资源指原本不存在,用户直接创造出来的数字资源,如直接的数字文本、数字摄像或数字录音文件等,这些资源通常没有相似等价物,与那些原物经数字化转换形成的相似等价物完全不同^[8,9]。常见的电子图书、电子期刊、网页、博客、科学数据、数据集、在线数据库、数字教育资源如教学课件和教学视听资料、数字图片、医学图像、数字遗产、数字全息图像、数字化X射线或放射线图像、数字证据、数字音频和视频、数字动画、数字媒体、数字图表、数字档案和数字化出版物(如数字化图书、期刊和报纸)等都属于数字资源。

数字资源保存是为了将来能够使用已创造或正创造的一些重要且富有价值的数字资源,而通过数字保存策略、措施、技术、方法和设备将其存储起来,以备将来之需。数字保存联盟(Digital

Preservation Coalition)认为数字资源保存是为了保证数字资源在需要时能够被持续获取而进行的一系列管理活动如资源规划、资源配置和技术使用等。这些数字资源可能是一个组织日常商业运营中所创造的记录,为特殊目的而创造的原生数字资源如教学资源,或数字化工程的产品等。数字保存联盟将数字资源保存分为长期保存、中期保存和短期保存。长期保存是指人们能够永久而持续地获取所需数字资源或至少数字资源中所包含的信息。中期保存是指人们在技术更新换代之后的一段时期内仍能继续获取所需的数字资源。短期保存是指在数字资源是否使用能够被预测的时期周期内,保证人们获取到数字资源,该周期不会延伸到可预测的使用周期之后太久和数字资源因技术变化而变得不可获取之后。美国图书馆联盟(ALA, American Library Association)将数字资源保存定义为:随着时间的流逝,为确保人们获取原生的或数字化的数字内容而采取的保存政策、策略和行动方案,其目标是随着时间的流逝,能确保已认证数字内容得到最真实和最精确的翻译和展现。其中数字保存政策指一个组织为未来使用而保存数字内容的承诺,详细说明将被保存内容的文件格式和保存级别,确保保存的数字信息能与数字信息标准一致;数字保存策略和行动方案主要指数字内容创造、整合和维护的策略和方案,如数字内容创造包括保存技术规范的清晰而完整的说明、元数据的结构化描述和管理、主文件的创造和过程的详细质量控制等;数字内容整合包括所有政策、策略和程序的文档化,所有数字对象

起源和变化历史的记录,认证机制等;内容维护包括强健的计算和网络基础设施、多个站点文件的存储和同步、文件的持续监测和管理、灾难阻止和恢复计划的创造和测试、政策和程序的定期评估和更新等^[10]。欧洲数字保存机构(DPE, Digital Preservation Europe)认为数字资源保存是为确保数字对象在将来能够被查找、翻译、使用和理解而采取的一系列行动,包括管理对象名称和地址、更新存储媒介、数字内容存档和跟踪硬件和软件变化,以确保数字对象在名称和地址、存储媒介及软硬件发生改变时仍能被打开和理解^[11]。国际空间数据咨询委员会(CCSDS, Consultative Committee for Space Data Systems)认为数字保存是在很长的时期内,以一个正确的和独立可理解的格式保存和维护信息的行为^[12]。英国研究委员会(Research Councils UK)认为数字保存是关于数字或电子对象维护和策划(包括选择、保存、维护、收集和存档)的所有活动^[11]。从数字保存国际组织对数字资源保存概念的定义可以看出,数字资源保存概念具有以下特征:①可读性,即可以查找、翻译、使用和理解等;②持久性,即保存时间的永久性;③真实性,内容完整与可靠;④动态性,保存策略随信息技术更新而发展变化;④准确性,已认证的数字内容能够得到最精确的翻译和展示;⑤功能性,保存的数字内容应不失其功能性,能够满足用户的功能性需求;⑥集成性,数字资源保存需要综合考虑保存的政策、策略、技术、设备、方法、措施、手段和一系列保存管理活动和行动。我们认为数字资源保存是保存人类原生或数字化

知识和文化遗产的活动,主要包括制定保存计划、政策、策略和方案;选择保存内容和元数据标准;建立保存的信息基础设施;提出保存方法、措施和手段;开发保存设备和技术;设定保存周期;执行保存计划和方案;实施保存项目;评估保存成本和效益;等等。

3 数字资源保存的发展史,研究现状

国外数字资源保存研究最早起源于英国学者罗德·唐纳德逊(Lord Donaldson)在1977年发表的一篇名为“信息传播和保存(*The dissemination and preservation of information*)”的文章^[13]。不过美国康奈尔大学图书馆副馆长安妮·R·肯尼雅(Anne R. Kenney)和林尼·K·皮尔逊尤尼(Lynne K. Personius)在1991年“美国情报学协会中期会议:多媒体研究”上发表一篇名为“数字保存——一项联合研究(*Digital Preservation—Joint Study*)”的文章^[14],才正式拉开了国外数字保存研究的序幕。至今(1992-2011),国外学者已经发表了1,500多篇数字保存方面的论文^③。研究主题涉及:数字保存的策略、模型、实现框架、机制、方法、实践和应用研究,数字保存项目成本效益的测度和评估,各国的数字保存项目研究,长期或大规模数字保存研究,数字保存的协同研究环境,分布式数字保存的协作途径,科学数据的数字保存,科学社团的数字保存需要研究等。论文中的大部分发表在计算机和图书情报档案方面的国际期刊如“Lecture Notes

in Computer Science(LNCS)”、“Library Trends”、“Library Collections, Acquisitions, and Technical Services”、“Journal of Documentation”、“American Society for Information Science and Technology”和“International Journal of Digital Curation”等,不过专注于数字保存主题的专业期刊较少,目前仅见“International Journal of Digital Curation”一种。除了发表大量论文外,国外学者还出版了大量相关论著,如澳大利亚学者罗斯·哈维(Ross Harvey)在2010年出版的《数字馆藏管理:如何做的手册》(*Digital Curation: A How-To-Do-It Manual*),英国学者玛里琳·迪安(Marilyn Deegan)和西蒙·蒂安(Simon Tanner)在2006年出版的《数字保存》(*Digital Preservation*),巴里·玛虹(Barry Mahon)在2003年出版的《数字保存》(*Digital Preservation*)等,夯实了数字保存研究的基础。另外,国外有很多国际性组织,如数字保存联盟(DPC)、欧洲数字保存机构(DPE)、数字保存中心(PrestoCentre)、国际互联网保存联盟(IIPC, International Internet Preservation Consortium)、欧洲保存与获取委员会(ECPA, European Commission on Preservation and Access)、图书馆联盟和机构的国际同盟(IFLA, International Federation of Library Associations and Institutions)、联合国教科文组织、国际空间数据咨询委员会、美国国会图书馆、各国图书馆和档案馆及它们之间的联盟等经常关注、支持和开

展数字保存方面的研究。这些国际性组织举办了很多数字保存方面的会议,如美国国会图书馆在2007-2011年举办的系列数字保存方面的会议,如2009年举办的“Designing Storage Architectures for Digital Preservation”和2011年举办的“Designing Storage Architectures for Preservation Collections”;联合国教科文组织发起的数字保存方面的系列会议,如2011举办的“International conference on preservation of digital information”和2012年将举办的“The Memory of the World in the Digital age: Digitization and Preservation”;2011年在德国举办的“数字保存峰会(Digital Preservation Summit)”;自2004年已举办8届的“数字对象保存国际会议(International Conference on Preservation of Digital Objects)”等。极大地推动了国外数字保存的发展。

国内数字保存研究起源于1997年南京大学信息管理学院吕斌发表的“数字信息如何保存?——电子信息时代的新挑战”文章。武汉大学信息管理学院刘家真和司莉在1999分别发表的“保护数字信息的长期存取策略”和“数字资源保存面临的挑战与对策”两篇文章,正式拉开了数字保存研究的序幕。至今(1997-2011),国内学者发表了390多篇中文论文,其中大部分论文发表在图书情报档案方面的期刊和一些大学学报。论文中的大部研究主题与国外类似。在书籍出版方面,国内数字保存方面的书籍相对较少,主要有两本:河北大学管理

^③ 数据库:SCI Web of Science;检索式:TI=(digital OR information) AND TI=(preservation OR conservation OR storage)

学院宛玲在2006年出版的《数字资源长期保存的管理机制》和北京联合大学谢永宪在2011年出版的《数字资源长期保存研究》。另外,国内举办的数字保存方面的会议相对较少,主要有2004和2007年在北京举办的“数字对象保存国际会议”和2004年在北京举办的“中欧数字资源长期保存国际研讨会”。

4 数字资源保存的研究进展分析与研究热点探测

主要利用陈超美开发的CitespaceII可视化软件及其提出的知识图谱绘制方法^[16-18]来分析数字资源保存的研究进展和探测它的研究热点与前沿。

4.1 数字资源保存的研究进展分析

通过构建学科知识基础(指引

文)随时间推移不断突变和演化的知识图谱和绘制学科发展趋势图,来分析数字资源保存的研究进展。

知识图谱的生成方法:首先用检索式“TI=(digital OR information) AND TI=(preservation OR conservation OR storage)”于2012年5月24日检索Web of Science(包括SCI-EXPANDED、SSCI、A&HCI、CPCI-S和CPCI-SSH索引数据库),得到1992年至2011年间的1,562篇文献,共52,632条引文数据(其中1,556篇文献,共52,558条引文数据是有效的)。然后设置软件的运行参数,如表1所示。

最后运行程序,执行寻找聚类(Find Cluster)和施引文献标题术语标识聚类主题两大功能,获得引文的共被引聚类知识图谱,分别采用时间序列(Time line)和时区(Time zone)的视图布局,绘制出引文共被引聚类的时间演化图谱和

时区演化图谱,如图1和图2所示。知识图谱的模块化(Modularity)系数接近0.88,说明图谱的网络结构非常好,形成了相互独立、边界清晰的聚类,聚类内部连接比较紧密。图谱的Mean Silhouette系数约为0.93,说明聚类视图比较可靠、清晰和直观,易于解读^[19,20]。图谱中红色圆圈表示爆发文献(指被引突增的文献),圆圈大小表示被引频次的高低,文献节点之间的连线表示两个文献之间的共被引关系,线的粗细表示两个文献之间的共被引频次或强度,线的长度表示两个文献共被引的时间跨度。蓝色字体是基于TF*IDF算法从聚类成员的施引文献标题术语中获得的聚类标签。

从图1和图2,我们发现:① 数字资源保存领域平均每篇文献约有21篇引文,说明国际上该领域的研究已经非常深入,有足够多的文献可供引用。② 知识图谱中出现549个被引频次高于1的引文(平均被引频次约为3次),1,045个共被引链接,形成了90个相互独立、边界清晰的共被引聚类(成员5个以上的聚类有42个),其中316篇引文和45个共被引聚类集中于1992—2005年间,2006到2010年间,也出现了一些被引频次大于1的引文(42篇)和共被引聚类(4个),说明1992年至2010年间,数字资源保存研究领域已经形成一些稳定集中的研究热点。③ 随着时间的推移,数字资源保存研究领域的一些研究主题得到更大程度的关注,如分别在1994、1997和2002年出现3篇被引突增(图中红色圆圈所示)的文献。被引突增的文献信息如表2所示。

表1 软件运行的参数名称及参数值

参数名称	参数值
项目及数据的路径	略
时间分割 (Time Slicing)	时间范围: 1992—2011; 分割长度: 1年
术语来源 (Term Source)	施引文献的题名、摘要、主题词和标引词
术语类型 (Term Type)	无 (None)
网络节点类型 (Node Types)	被引文献 (Cited Reference)
连接强度 (Links Strength)	Cosine
Top N per slice ^④	30
网络约简 (Pruning)	网络关键路径发现 (Pathfinder) 及约简 (Pruning Sliced Networks) 算法 ^⑤
可视化 (Visualization)	静态聚类视图 (Cluster View-Static) 和融合网络 (Merged Network)
术语和文献标签	按中心度 (Centrality) 标注

^④ “Top N per slice” 指每个时间周期或分割内被引频次最高或出现频次最高的文献或术语。
^⑤ 该算法能够很好地保持网络内聚性和刻画网络如何演化的本质。参见文献: CHEN C, MORRIS S. Visualizing Evolving Networks: Minimum Spanning Trees versus Pathfinder Networks [C]// Proceeding of the Ninth annual IEEE conference on Information visualization, IEEE Computer Society, Washington DC, USA, 2003: 67—74.

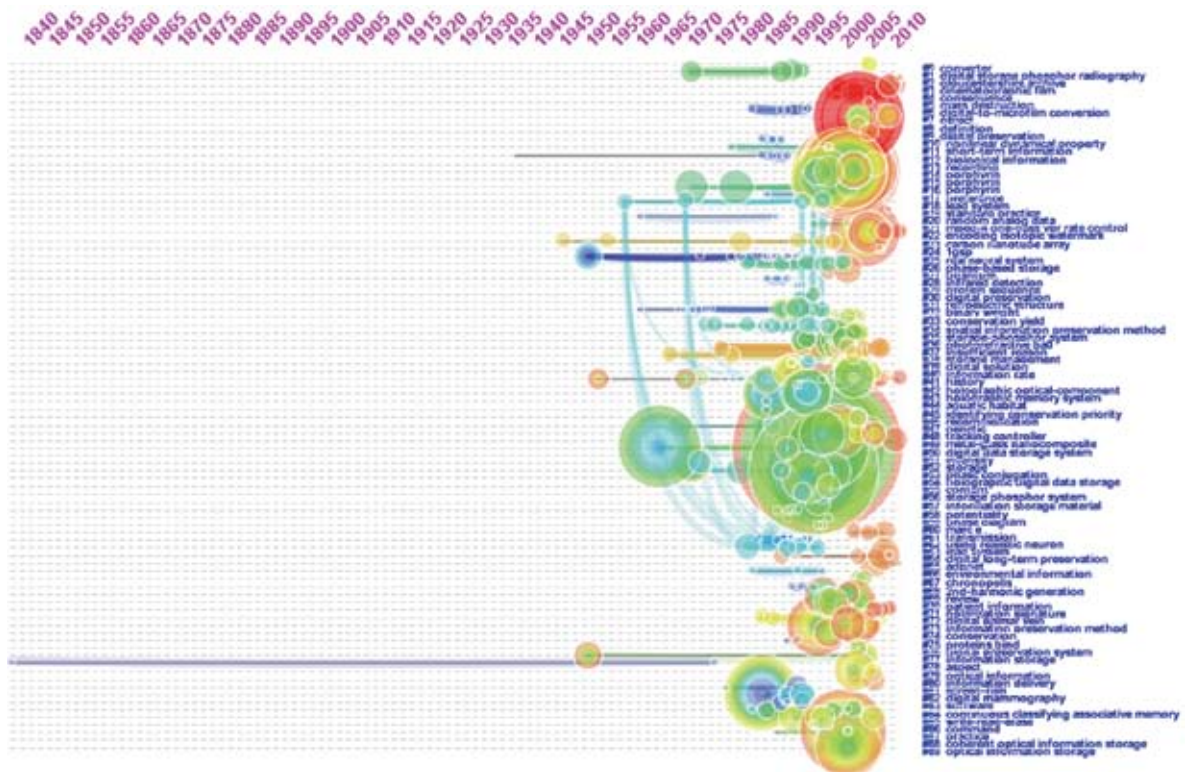


图1 引文共被引聚类的时间演化图谱 (在表1所设置的参数条件下, 有1,194篇有效施引文献, 共24,763条有效引文, 无效引文有74条, 融合网络的节点有549个, 链接1,045个)

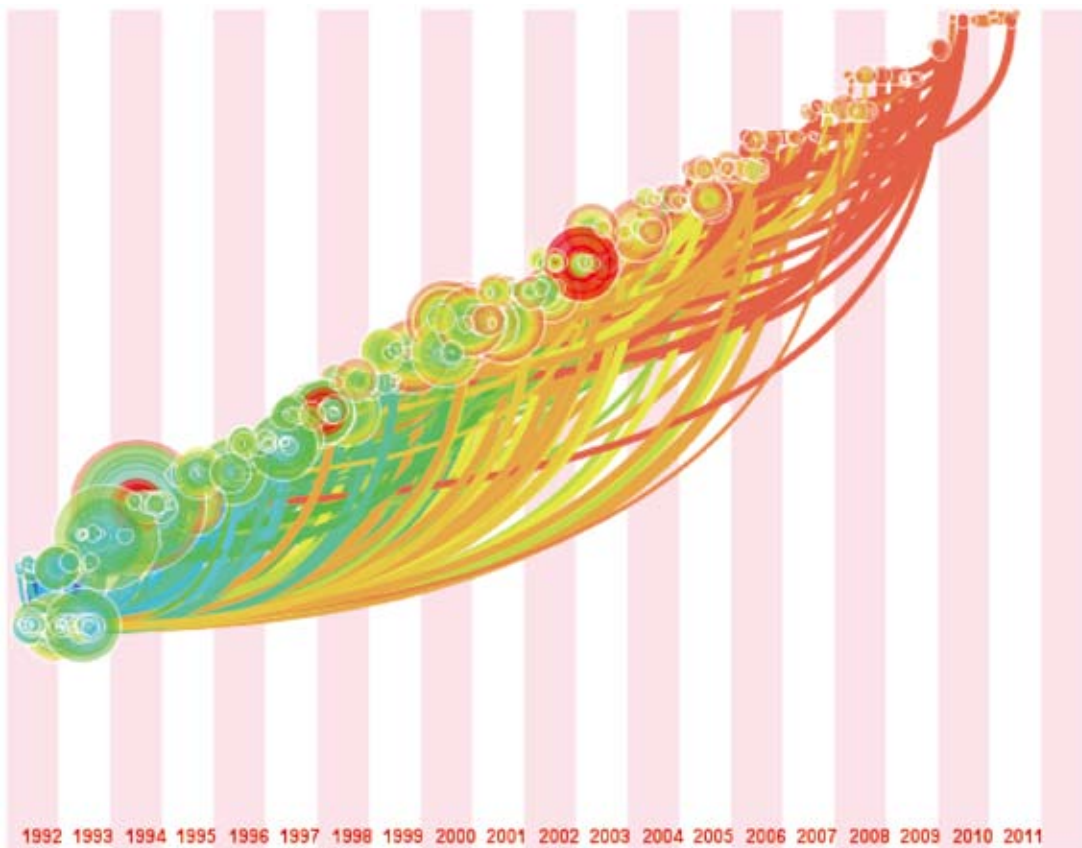


图2 引文共被引聚类的时区 (Time Zone) 演化图谱

表2 引文的篇名、最初被引年代、突增强度 (Strength)、被引突增的起始和结束年代、时间条

篇名	年代	强度	开始	结束	1992-2011
Volume Holographic Storage and Retrieval of Digital Data	1994	4.6197	1995	1999	
Pixel-matched holographic data storage with megabit pages	1997	3.5317	1998	2003	
Reference Model for an Open Archival Information System (OAIS)	2002	3.7408	2008	2011	

其中前2篇文献的主题是全息存储^[21,22],它们在1995至2003年期间被引突增,被此期间发表的31篇数字资源保存文献引用了31次,施引文献的主题也是全息存储,研究者大部分属于物理学、电子学、材料科学、信息科学和光学学科。第3篇文献的主题是“开放文档信息系统参考模型”^[12],被2008至2011年间发表的12篇数字资源保存文献引用了12次,施引文献的主题大部分是数字长期保存方面的,涉及数字长期保存模型^[23]、数字保存对象的描述语言(如XML)^[24]、数字长期保存实践^[25](相同篇名文章有6篇)和新兴保存标准^[26]、数字长期保存工具和服务^[27]等。研究者大部分属于图书情报档案管理学学科。从爆发文献主题及其时间演化来看,全息存储和数字资源长期保存是数字资

源保存的研究热点,其中数字资源长期保存是自2002年开始,刚兴起的研究前沿。

为从数量和质量上反映数字资源保存领域的发展趋势,我们绘制了数字资源保存研究文献在1992-2011年间的数量分布图,并将图1或图2所示中心度较高的关键节点文献和爆发文献按时间标示在数量分布图中,从而获得质与量相结合的学科发展趋势图,如图3所示。这些关键节点文献或爆发文献能够作为发展阶段划分的重要标识,可视为一个阶段内学科知识发展的小突破^[28]。

从图3可以看出,在1993年、1994年、1997年、2000年和2002年分别出现一篇中心度较高或被引突增的文献。1993年、1994年、1997年和2000年出现的4篇文献主

要关于全息存储研究,在这4篇文献的刺激和推动下,数字资源保存研究不断向前发展,分别在1994年、1997年和2001年出现三次发展高峰。2001年之后,经过一个小幅度下滑之后,在2002年,空间数据系统咨询委员会发布了名为“开放文档信息系统参考模型”的蓝皮书^[12],奠定了数字资源长期保存研究的基础,促使“数字资源长期保存”成为2002年以来新兴的研究前沿和热点(比如,2002年至2011年期间,国外学者发表了63篇数字资源长期保存方面的文献,约占1992年至2011年期间所有数字资源长期保存文献数量[91篇]的69%^①)。在数字资源长期保存研究主题的加入和推动下,数字资源保存在2005年达到发展高峰后,然后一直保持高速平衡的发展趋势。另外,通过查看数字资源保存研究文献的标题和摘要,我们发现数字资源保存研究正从2003年之前注重保存材料、设备、工艺和技术的理工科研究范畴(研究重点是全息存储和光存储)向同时注重理工科研究范畴和数字长期保存规划、策略、实践、方法和模型的人文学科研究范畴方向发展。

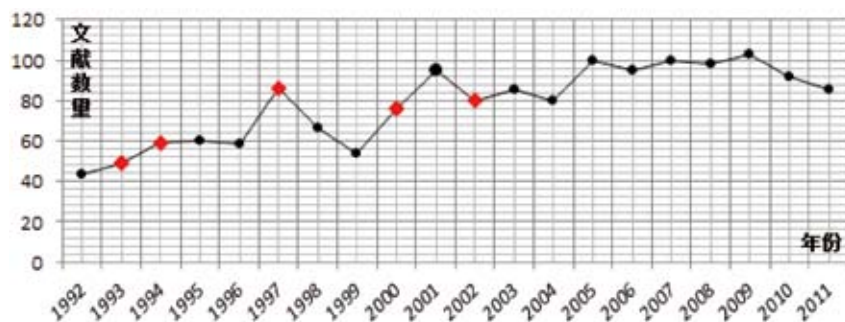


图3 质与量相结合的数字资源保存发展趋势图

① 数据库=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, CCR-EXPANDED; 检索式: TI=(digital OR information) AND TI=("long term" OR "long-term" OR "long time" OR "long periods") AND TI=(preservation OR conservation OR storage OR Memory)

表3 热点研究主题的名称、出现年代区间、共现频次和中介中心度的总和

主题名称	出现年代区间	共现频次总和	中心度总和
数字全息存储 (digital holographic storage)	1993-2007	89	0.15
长期保存 (long-term preservation)	1995-2010	81	0.08
系统 (system)	1992-2006	81	0.12
光存储 (optical storage)	1992-1999	75	0.11
数字图书馆 (digital library)	1996-2008	59	0.06
数字图像存储 (digital image storage)	1994-1998	52	0.13
数字存储技术 (storage technology)	1993-2010	49	0.22
模型 (model)	1997-2004	38	0.08
地理信息系统 (geographic information system)	1992-2006	37	0.07
数字存储荧光体 (digital storage phosphor)	1992-2003	36	0.03
数字存储示波器 (digital storage oscilloscopes)	1994-2003	22	
数字存储设备 (storage devices)	1998	17	0.01
国家数字信息基础设施 (national digital information infrastructure)	2009	10	
数字保存策略 (digital preservation strategy)	2007	7	
数字保存计划 (preservation planning)	2000-2008	13	0.01
数字保存行动 (Action)	2010	10	
数字保存项目 (preservation program)	2009-2010	15	
数字保存安全 (security)	2009	8	
数字保存方案 (scheme)	2009	3	
闪存存储 (flash memory)	2010	5	

从表3研究主题的共现频次或中介中心度总和来看,数字全息存储、长期保存、系统、光存储、数字图书馆、数字图像存储、数字存储技术、模型、地理信息系统、数字存储荧光体和数字存储示波器是数字资源保存研究领域的研究热点。不过从主题出现的年代区间来看,国家数字信息基础设施、数字保存策略、数字保存计划、数字保存行动、数字保存项目、数字保存安全、数字保存方案和闪存存储是数字资源保存研究领域中新兴的研究前沿。

上述方法虽然能够反映数字资源保存的研究热点,并且在一定程度上能够看出数字资源保存的近

期研究主题,不过如果不列出那些共现频次或中心度总和较少的主题,就无法明确看出数字资源保存近期的研究前沿。因此我们利用J. Kleinberg和陈超美提出的爆发词探测算法和技术^[15,31]从施引文献标题和摘要分词所得的名词术语集中获得能够反映研究前沿的爆发词或突变词,如表4所示。

从表4爆发词开始爆发的时间和爆发结束的时间可以看出:

① 数字存储 (digital-storage) 术语在2007之前使用较多,而数字保存 (digital-preservation) 术语在2007年之后使用较多,从“数字存储”术语到“数字保存”术语的转

变,反映出数字资源保存研究方向和理念的转变; storage的对象只是普通的物体,而preservation的对象可能是人类文明遗产。确实,联合国教科文组织非常强调“数字遗产”概念。② 国家数字信息基础设施、保存规划、数字对象、数字图书馆和数字数据是数字资源保存研究领域近几年新兴的研究前沿。

4.2.2 基于引文共被引聚类知识图谱的研究热点探测

引文共被引聚类知识图谱的生成方法和参数配置与图1一样,不过在结果视图的布局(Layout)上,我

表4 爆发词的名称、爆发权重 (weight)、爆发开始和结束的时间

名称	weight	begin	end
国家数字信息基础设施（national-digital-information-infrastructure）	4.06	2009	
保存规划（preservation-program）	3.61	2009	
数字对象（digital-object）	4.66	2008	
数字图书馆（digital-library）	4.28	2008	
数字数据（digital-data）	3.84	2007	
数字保存（digital-preservation）	7.16	2007	
数字存储（digital-storage）	4.12	1992	1996
数据存储（data-storage）	3.77	2005	2008
存储荧光体系统（storage-phosphor-system）	3.60	2002	2003
存储能力（storage-capacity）	3.65	1997	2000
高密度信息存储（high-density-information-storage）	3.55	1997	1997
数字全息数据存储（digital-holographic-data-storage）	4.10	1995	1999
数字存储示波器（digital-storage-oscilloscopes）	4.66	1994	2000
全息存储（holographic-storage）	4.05	1994	1998
数据存储（digital-storage-media）	3.27	1994	1997
存储荧光体（storage-phosphor）	3.12	1993	1997
光信息存储（optical-information-storage）	4.42	1992	2001

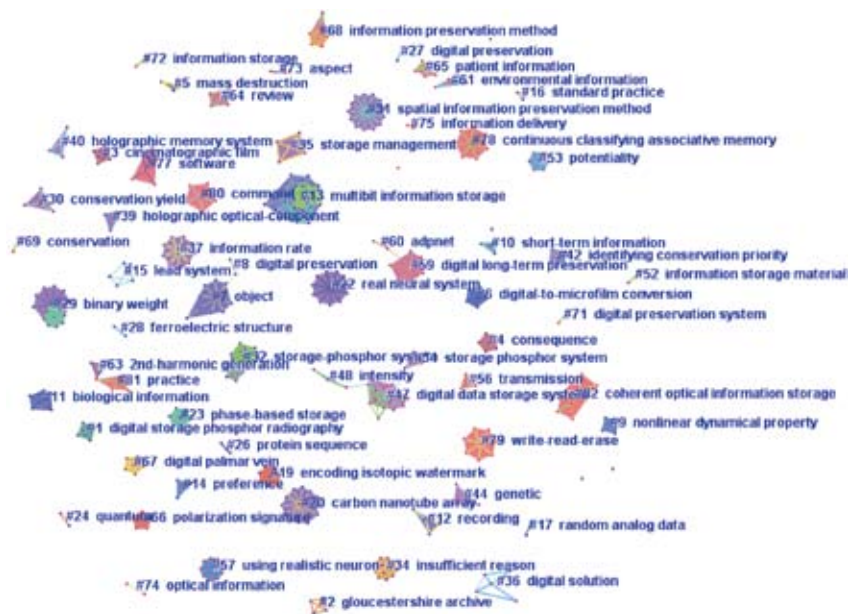


图5 TF*IDF权重方法+施引文献标题术语组合的共被引聚类知识图谱(图谱的Modularity值约0.9, Mean Silhouette值接近0.95, 说明聚类的效果非常好)

们采用聚类视图 (Cluster View)，然后分别使用TF*IDF、Log似然率 (LLR, Log-likelihood Ratio) 和互信息 (MI, Mutual Information) 三种术语权重方法分别从各个聚类引文对应施引文献的标题、标引或摘要术语集中选择权重最高的术语标引各个聚类的主题，共得到9种聚类主题术语不一样的聚类视图。其中TF*IDF权重方法+施引文献标题术语组合的共被引聚类知识图谱，如图5所示。

从图5可以看出, 数字长期保存(digital long-term preservation)、多位信息存储(multibit information storage)、信息率(information rate)、数字数据存储系统(digital data storage system)、数据保存对象(object)、数字保存实践(practice)、存储荧光体系统(storage-phosphor system)、光学信息存储(optical information storage)、空间信息保存方法(spatial information preservation method)和信息保存方法(information preservation method)是数字资源保存领域的研究热点。

通过不同的标引方法(三种权重方法*三种标引术语来源,共9种标引方法)标引的聚类视图,反映聚类主题的术语是不一样的。因此,为了统一和综合性地揭示数字资源保存领域的研究热点,我们首先从共被引聚类图谱的90个聚类中选择聚类成员和施引文献数量较多、年代较新的9个聚类,然后对反映聚类主题的9种不同术语集进行归纳总结和融合,得出9个聚类的编号、成员和施引文献数量,聚类成员的平均年份,反映聚类主题的Top主题术语,以及聚类的主要施引文献,如表5所示。

表5 9个聚类的编号、成员和施引文献数量、平均年份、Top主题术语及主要施引文献

聚类编号/成员数量/施引文献数量	平均年份	Top主题术语	主要施引文献
41/66/87	1992	数字保存历史 (history)、理论 (theory)、机制 (mechanism)、服务 (service)、未来 (Future)、稳定性 (stabilization) 和可持续性 (Sustainability)	图书情报档案管理领域学者对该主题研究较多 (具体参见脚注⑨)
14/19/8	2000	多位信息存储 (multibit information storage) 和基于分子的信息存储 (molecular-based information storage)	化学领域学者对该主题研究较多 (Gryko DT (2000) synthesis of "porphyrin-linker-thiol" molecules with diverse linkers for studies of molecular-based information storage; Gryko DT (2000) synthesis of thiol-derivatized ferrocene-porphyrins for studies of multibit information storage)
35/18/5	2000	存储荧光体系统 (storage-phosphor system) 和放射线数字影像 (radiography)	放射医学领域学者对该主题研究较多 (Goo JM (2000) digital chest radiography with a selenium-based flat-panel detector versus a storage phosphor system: comparison of soft-copy images)
48/13/9	1999	全息数字数据存储 (holographic digital data storage)、设计 (design)、误比特率 (bit-error rate) 和相联存储器的全息数字数据存储 (content-addressable holographic digital data storage)	物理学和光学领域学者对该主题研究较多 (Fengcheng Ma (2010) the construction of digital resources storage platform based on holographic technology; John R (2005) an input-data page modulation scheme for content-addressable holographic digital data storage; Heanue JF (1994) Volume Holographic Storage and Retrieval of Digital Data)
37/12/3	1998	成功 (success)、成本 (cost)、动态信息 (active information)、内生信息 (endogenous information)、特定问题信息 (problem-specific information) 和信息强度 (information density)	Dembksi William A. (2009) conservation of information in search: measuring the cost of success; Li Xin (2010) maximum-information storage system: concept, implementation and application
50/12/12	1992	数字数据存储系统 (digital data storage system)、全息光盘 (holographic disk)、超高转移率 (ultra-high transfer rate)、大容量 (high capacity)、快速获取 (fast access memory)、固态硬盘 (solid-state memory)、全息数据存储 (holographic data storage)。	物理学、光学和材料科学领域学者对该主题关注较多 (Orlov SS (2000) ultra-high transfer rate high capacity holographic disk digital data storage system; Ma JA (2000) ruggedized digital holographic data storage with fast access; John R (2005) holographic digital data storage using phase-modulated pixels)
7/8/5	2005	数字保存对象 (object)、保存模型 (model)、迁移 (migration)、保存计划 (preservation planning)、XML语言 (xml languages)、长寿 (longevity)、评估 (evaluation) 和实施 (execution)	信息科学和图书情报档案管理领域学者对该主题关注较多 (Becker Christoph (2008) systematic characterisation of objects in digital preservation: the extensible characterisation languages; Luan Feng (2010) quality requirements of migration metadata in long-term digital preservation systems; Caplan Priscilla (2010) the florida digital archive and daitss: a model for digital preservation)
64/7/7	2003	数字长期保存 (digital long-term preservation)、数字图书馆 (digital library)、数字保存研究 (digital preservation research)、真实性 (authenticity)、完整性 (integrity)、用户研究 (user studies)、安全模型 (security model)、安全方面 (security aspect) 和长期存储 (long-term storage)	图书情报档案管理领域学者对该主题关注较多 (Chowdhury Gobinda (2010) from digital libraries to digital preservation research: the importance of users and context; Schott Maik (2010) ensuring integrity and authenticity for images in digital long-term preservation; Schott Maik (2010) extending the clark-wilson security model for digital long-term preservation use-cases; Voges Christoph (2011) long-term storage of digital data on cinematographic film)
87/5/4	1996	数字保存的技术与实践 (art and practice)	信息科学和图书情报档案管理领域学者对该主题关注较多 (Lee KH (2002) the state of the art and practice in digital preservation; Muir A (2001) legal deposit and preservation of digital publications: a review of research and development activity)

⑨ 1. Hamid S (1998). Constructing a global time capsule: challenges in the digital preservation of society's cultural memory archive.

从表5可以看出,近年来,国外学者除了从技术视角继续围绕数字资源保存的传统话题(如多位信息存储、分子信息存储、存储荧光体系、放射线数字影像存储、全息数字数据存储、数字数据存储系统、全息光盘、超高转移率、固态硬盘和数字存储技术等)进行研究之外,越来越多的学者开始从社会科学和人文学科视角围绕数字资源保存开展研究,涉及数字资源保存的历史、理论、机制、服务、未来、稳定性和可持续发展;保存的成本、对象、模型、计划、真实性、完整性、评估和实施;数字资源寿命、长期保存和数字图书馆;用户研究、保存实践和安全等。

5 数字资源长期保存研究进展综述

数字资源长期保存是数字资源保存研究的重点,也是亟待解决的难点。目前国内外图书情报和信息科学学科的学者和机构对数字资源长期保存关注较多,很多数字资源保存方面的实践项目主要是为了解决数字资源长期保存的难题而提出的。因此笔者主要聚焦于数字资源的长期保存,从研究主题、最佳实践、保存周期、保存技术和方法、保存策略等几个重要方面解读数字资源长期保存的研究进展^[32-35]。这些方面也是目前国内外学者研究的前沿与热点。

5.1 数字资源长期保存的研究主题

北京师范大学田硕和黄国彬通过对近10年(2000-2009)国外数字资源长期保存的期刊论文进行归纳总结,发现:国外数字资源长期保存研究表现出理论研究和实践研究相结合、技术研究和非技术研究并重的特点,其研究主题以长期保存的相关技术和保存策略及保存项目为主,并注重对具体问题、具体保存对象的研究,其研究领域已经逐渐渗透到社会生活的各个层面。此外,国外数字资源的长期保存研究还涉及保存的战略与基础设施建设、保存成本管理、元数据研究、相关认证与评估、教育与培训以及知识产权管理等。

5.2 数字资源长期保存的最佳实践

数字保存研究的发展离不开实践,也离不开数字保存理论研究的指导。两者是辩证统一的关系,相辅相成,缺一不可。因此各国相关机构和组织在数字资源长期保存理论探索成果的指导下,纷纷开展数字资源长期保存方面的实践活动,发展了很多具体成熟、规模较大、应用广泛、影响较广、经济和社会效益突出的数字资源长期保存项目。如美国斯坦福大学图书馆在1999年发起的,用于实时收集和保

存在线数字资源如在线图书和期刊等的LOCKSS项目;美国国会图书馆在2002年通过立法确立的,用于收集和保存各类数字资源如Web信息、视频、音频、数字期刊、电子图书、数字电视等的国家数字信息基础设施和保存项目:NDIIPP(National Digital Information Infrastructure Preservation Program);英国高等及继续教育基金会联合信息系统委员会JISC在1998资助的,致力于研究数字资源长期保存的策略、方法和实际问题,为图书馆最佳数字保存实践提供指导的高校研究型图书馆联盟的数字样书保存项目:CEDARS(CURL Exemplars in Digital Archives);由荷兰国家图书馆牵头,欧洲7个国家图书馆和3个主要出版社共同参与,于1998年启动的网络化欧洲存储图书馆项目:NEDLIB(Networked European Deposit Library),该项目主要致力于研究长期保存基础结构和欧洲存储图书馆机制,以保证电子出版物的长期获取;德国教育和研究委员会(German Federal Ministry of Education and Research)在2003年资助的数字资源长期保存专家网络项目:NESTOR(the Network of Expertise in Long-term Storage of Digital Resources);澳大利亚国家图书馆在1995年发起的澳大利亚网络文档资源的保存和获取项目:PANDORA(Preserving

2. Smith A (2000). Going digital: strategies for access, preservation and conversion of collections to a digital format.
3. Daniels T (2001). A history of information storage and retrieval.
4. Day M (2003). Digital preservation and metadata: history, theory, practice.
5. Hockx-Yu Helen (2006). Digital preservation in the context of institutional repositories.
6. Horimai Hideyoshi (2007) Holographic information storage system: today and future.
7. Barateiro Jose (2008). Digital preservation of scientific data.
8. Barateiro Jose (2008). Using a grid for digital preservation.
9. Conway Paul (2010). Preservation in the age of google: digitization, digital preservation, and dilemmas.
10. Walters Tyler O (2010). Economics, sustainability, and the cooperative model in digital preservation.

and Accessing Networked Documentary Resources of Australia); 中国国家图书馆在1995年启动的网络信息资源采集与保存项目: WICP (Web Information Collection and Preservation); 北京大学在国家“973”和“985”项目支持下,于2002年启动的中国Web信息博物馆项目,并开发了一款用于历史网页存盘和回放的中国历史网页信息存储与展示系统。

5.3 数字资源长期保存周期

数字资源保存周期长短的确定是数字资源保存研究领域中的一个非常重要但又难以解决的问题,如果保存周期过长,会导致保存成本过高,如果保存周期过短,可能会因技术更新换代、文件损坏和载体发生故障而损失一些重要而有价值的资源,给用户带来巨大损失。因此保存周期应综合考虑数字资源的类型、生命周期、质量和价值、重要程度及未来使用的概率等因素来确定,以使保存成本最小化,保存利益最大化。目前国内外相关机构和学者关注比较多的是数字资源长期保存的周期,对该周期的确定,不同学者和机构的认识不尽相同。国际图联外联与服务部主任伯格·沃赫尔(Ingeborg Verheul)认为这个周期可能是5年或更长^[36]。国际空间数据咨询委员会认为保存周期应该长到无需考虑因技术和软硬件环境变化、载体故障、文件损坏、数据格式变化及用户社区变化而导致的数字资源获取困难或失败问题,并且这个期限可以拓展到无限期的将来^[11]。英国研究委员会(Research

Councils UK)认为任何项目的数据保存和获取周期应该不少于10年,其中临床或主要社会、环境和重要遗产项目的数据,应该保留20年,当然能够永久保存或按资助者的数据政策中要求的保存周期保存更好^[37]。上田修一(1999)认为:目前电子媒体寿命短,一般只有10年左右,因此电子信息的长期保存必须考虑这一因素,在技术上应开发出寿命达到100年以上的媒体^[38]。

5.4 数字资源长期保存的相关技术和方法

数字资源长期保存技术是数字资源长期保存的重要支撑,是国内外学者研究的重点和热点。综观国内外学者在数字资源长期保存技术方面的研究发现:国内外学者除了对数字资源标准化技术、迁移和仿真技术、封装技术、更新技术、数据恢复和再造技术、比特流拷贝即数据备份技术、复制技术等传统保存技术感兴趣之外,更对新兴的基于云计算的动态存储和获取技术、元数据保存技术、网格存储与获取技术、分布式数字保存网络或系统技术、海量数字资源保存技术、机构仓储技术、数字资源长期保存技术、虚拟存储技术、文化遗产的数字保存技术如数字博物馆、网络信息资源如Web2.0内容及博客内容的采集与保存技术、网络关联数据的格式保存以及保存系统、共享技术和互操作技术等感兴趣。

5.5 数字资源长期保存的策略

通过梳理国内外相关研究文献,发现目前数字资源长期保存的

策略主要有:(1) 合作保存策略。数字资源长期保存项目是一个非常巨大的工程,单凭个别机构或图书馆无法完成,需要众多机构或图书馆的参与。上述的数字资源保存联盟、国际性组织和很多大型数字资源长期保存项目基本上都是采取合作策略;(2) 技术保存策略。技术是数字资源保存项目得以实现的基本保障,好的技术可以使数字资源长期保存项目的效果成倍地放大。不同的数字资源适用的保存技术不尽相同,因此针对不同类型的数字资源和保存要求,科学合理选择相应的技术方案非常重要。目前常用的技术保存策略主要有:基于对等网的分布式保存和开放源代码的保存策略、为避免软硬件变化导致数字资源获取失败的仿真、封装和迁移技术保存策略、可以保证海量数字资源实时保存和获取的云存储策略等。(3) 风险管理策略。数字资源长期保存是一个耗时很长、动态变化的工程,中间难免出现一些风险,因此需要考虑数字资源长期保存过程中可能出现的风险,制定相应的风险策略。目前国外数字保存实践采取的风险策略主要有:通过合作开发和建立联盟的方式,以降低保存成本,提高保存效果;通过建立保存网络和构建责任体系的方式,以建立风险发生时的快速反应和处理机制;通过建立数据定期备份机制,以避免数字资源丢失的风险;通过应用仿真、封装和迁移技术,以避免技术更新导致的数字资源获取失败的风险;通过资金预算风险管理,以避免资金断链或不足。(4) 经济策略。数字资源长期保存是一个复杂的系统性工程,需要持续的高成本的技术与管理投入,所以经费问题是数字资源长期保存

需要重点考虑的问题。国外数字资源长期保存的经济策略主要有：获取国家和地方政府的研究与开发经费支持、各参与方共同分担、建立专门的数字资源保存研发基金、数字保存资金预算管理和周期成本计算、国家或政府机构专项投资、寻找赞助商等。

6 结语

通过对数字资源保存方面的1,562篇文献及52,632条引文数据进行时间序列分析、术语与关键词共现网络和引文共被引聚类网络的知识图谱分析,我们发现:

(1) 国际上数字资源保存相关

研究已经非常深入,有足够多的文献可供引用(按总引文数与总文献数之比来测算,平均每篇文献接近31篇引文)。

(2) 1992年至2010年间,数字资源保存已经形成夯实的知识基础(5万多条引文)和一些稳定集中的研究热点(形成了90个聚类),有一定的学术影响力(平均每篇文献被引6.88次),正稳步向前发展。

(3) 数字资源保存有两条研究主线:一是物理学、光学、电子学、放射医学、材料科学和信息科学等理工科学者从技术视角研究数字全息存储、光存储、数字图像存储、数字存储荧光体、数字存储示波器和闪存存储等。二是图书情报档案管

理学科学者从信息技术和人文相结合的视角研究数字图书馆,数字资源长期保存及相关的系统、模型、技术、方法、策略、计划、实践、行动、项目、安全、方案和国家数字信息基础设施等,这些主题也是近年来刚兴起的研究前沿与热点。

(4) 数字全息存储和光存储是1992年至2011年期间的持续研究热点,如1993—2011年期间,共发表557篇“数字全息存储”主题的文章(平均被引频次8.37次和平均每年发表28篇论文)^⑩和788篇“光存储”主题的文章(平均被引频次13.69次和平均每年发表39篇论文)^⑪。而“数字长期保存”是自2002年开始兴起的研究热点与前沿主题。

(5) 通过国家共现网络图谱分析(如附图6所示),我们发现,数字资源保存研究实力比较强的国家有:美国(出现频次为460次)、中国(124次)、德国(88次)、英国(76次)和日本(61次),不过美国和中国的中介中心度较低,都为0,说明尽管美国和中国的研究实力比较强,但与其他国家之间交流非常少,其中美国与加拿大交流得比较多,而中国与比利时(Belgium)交流得比较多。研究实力比较强,且与其他国家交流比较多的国家有:德国(中心度为0.22)、日本(0.08)和英国(0.06)。西班牙(频次为28)和瑞士(频次为13)的研究实力虽然不是太强,不过他们与其它国家之间的交流也比较多(中心度分别是0.16和0.1)。

(6) 通过机构共现网络图谱分析,我们发现美国北卡罗来纳州立

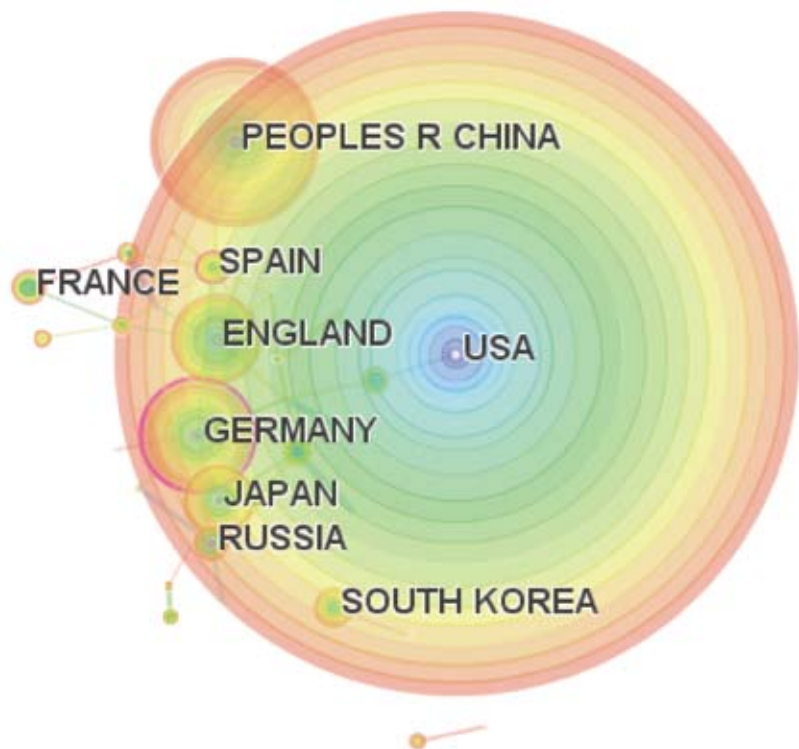


图6 国家共现网络知识图谱分析

^⑩ 数据库=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, CCR-EXPANDED; 检索式: TI=(digital OR information OR data OR image) AND TI=(holographic OR holography) AND TI=(preservation OR conservation OR storage OR Memory)

^⑪ 数据库=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, CCR-EXPANDED; 检索式: TI=(digital OR information OR data OR image) AND TI=(optical) AND TI=(preservation OR conservation OR storage OR Memory)

大学（出现频次为25次）、加州大学河滨分校（21次）、斯坦福大学（11次）和IBM公司（11次），以及中国科学院（18次）的研究实力比较强。

（7）在标引共被引聚类主题时，我们发现：在标引效果上，标题术语+TF*IDF组合方法的标引效果最

好，选择的术语能够全面、准确地反映聚类主题；在聚类主题术语来源上，标题术语集和标引术语集的主题相关性最高；术语选择方法上，TF*IDF权重方法最好，其次是Log似然率，互信息方法经常会选出很多大而空，不相关的术语，不能解读具体的研究热点与前沿主题。

（8）我们融合术语和关键词共现网络，以及引文共被引聚类网络来识别数字资源保存研究领域的热点与前沿，可以有效克服单独使用引文共被引聚类网络识别热点与前沿的时滞性问题（网络选择的引文需要达到一定频次，而达到一定频次需要一定时间）。

参考文献

- [1] LOPEZ-FITZSIMMONS B. Digital preservation and metadata: history, theory, practice [J]. Library Resources & Technical Services, 2003, 47(3): 135-136.
- [2] DAY M. Digital preservation and metadata: history, theory, practice [J]. Journal of the Society of Archivists, 2003, 24(2): 243-244.
- [3] TERMENS M. Research and development in digital preservation: an international review [J]. Profesional De La Información, 2009, 18(6): 613-624.
- [4] 刘家真.我国数字资源保存状况调查[J].中国图书馆学报,2006(5):71-75.
- [5] 吴振新,刘建华,张玫,等.数字资源长期保存:当前进展和最佳实践 -- 2007年数字资源长期保存国际会议(iPRES2007)综述[J].现代图书情报技术,2007,157(11):1-6.
- [6] 李丹,向菁.协作与实践:数字资源长期保存工具及方法 -- 2008年数字资源长期保存国际会议(iPRES2008)综述[J].图书馆理论与实践,2009(11):70-72.
- [7] 梁战平.中国情报学百科全书[M].中国大百科全书出版社,2010:257.
- [8] 邹志仁.信息学概论[M].南京:南京大学出版社,2011:60-79.
- [9] DPC-Digital Preservation Coalition. Introduction - Definitions and Concepts [EB/OL]. [2012-07-20]. <http://www.dpconline.org/advice/preservationhandbook/introduction/definitions-and-concepts>.
- [10] ALA - American Library Association. Definitions of digital preservation[EB/OL]. [2012-07-20]. <http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.pdf>.
- [11] DPE - Digital Preservation Europe. What is Digital Preservation? [EB/OL]. [2012-07-20]. <http://www.digitalpreservationeurope.eu/what-is-digital-preservation/>.
- [12] CCSDS (Consultative Committee for Space Data Systems). Reference Model for an Open Archival Information System (OAIS) [M]. Blue Book, Issue 1. Washington, DC (US): CCSDS Secretariat, PP.1-11. <http://public.ccsds.org/publications/archive/650x0b1.PDF>.
- [13] DONALDSON L. The dissemination and preservation of information [J]. Aslib Proceedings, 1977, 29(8): 272-278.
- [14] KENNEY A R, PERSONIUS L K. Digital Preservation -- Joint Study [J]. Studies in Multimedia: State-Of-The-Art Solutions in Multimedia and Hypertext, 1992: 47-52.
- [15] CHEN C. CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature [J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377.
- [16] CHEN C M, IBEKWE-SANJUAN F, HOU J H. The Structure and Dynamics of Co-Citation Clusters: A Multiple-Perspective Co-Citation Analysis [J]. Journal of the American Society for Information Science and Technology, 2010, 61(7): 1386-1409.
- [17] CHEN C. Searching for intellectual turning points: Progressive knowledge domain visualization [J]. Proceedings of the National Academy of Sciences of the United States of America (PNAS), 2004, 101(1): 5303-5310.
- [18] 陈超美,陈悦,侯剑华.Cite Space II,科学文献中新趋势与新动态的识别与可视化[J].情报学报,2009,28(3):401-421.
- [19] NEWMAN M E J. Modularity and community structure in networks [J]. PNAS, 2006, 103(23): 8577-8582.
- [20] ROUSSEUW P J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis [J]. Journal of Computational and Applied Mathematics, 1987 (20): 53-65.
- [21] HEANUE J F, BASHAW M C, HESSELINK L. Volume Holographic Storage and Retrieval of Digital Data [J]. Science, 1994, 265(5173): 749-752.
- [22] SHELBY R M, HOFFNAGLE J A, BURR G W. Pixel-matched holographic data storage with megabit pages [J]. Optics Letters, 1997, 22(19): 1509-1511.
- [23] PRISCILLA C. The Florida Digital Archive and DAITSS: a model for digital preservation [J]. Library Hi Tech, 2010, 28(2): 224-234.
- [24] BECKER C, RAUBER A. Systematic Characterisation of Objects in Digital Preservation: The eXtensible Characterisation Languages [J]. Journal of Universal Computer Science, 2008, 14(18): 2936-2952.
- [25] HOWARD R I. Networking for Digital Preservation: Current Practice in 15 National Libraries [J]. Library Resources & Technical Services, 2008, 52(4): 275-276.
- [26] THOMAS H, JANET E, CORDIAL M A, et al. Developments in Digital Preservation at the University of Illinois: The Hub and Spoke Architecture for Supporting Repository Interoperability and Emerging Preservation Standards [J]. Library Trends, 2009, 57(3): 556-579.
- [27] JOSEPH J, SANGCHUL S. Robust Tools and Services for Long-Term Preservation of Digital Information [J]. Library Trends, 2009, 57(3): 580-594.
- [28] 荣兰,王续琨.信息经济学研究主题迁移的可视化分析[J].情报科学,2011,29(5):757-762.
- [29] BRANDES U. A faster algorithm for betweenness centrality [J]. Journal of Mathematical Sociology, 2001, 25(2): 163-177.
- [30] CHEN C. The centrality of pivotal points in the evolution of scientific networks [C]// Proceedings of the International Conference on Intelligent User Interfaces (IUI 2005), ACM Press, 2005: 98-105.
- [31] KLEINBERG J. Bursty and hierarchical structure in streams [C]// Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 2002: 91-101.
- [32] 田硕,黄国彬.近十年国外数字资源长期保存研究综述[J].图书馆杂志,2011,30(7):8-13.
- [33] 黄旭,毕强.国内外数字资源长期保存研究现状与进展[J].图书馆学研究,2009(1):25-28.
- [34] 贺妹玮.国外数字资源长期保存发展历程[J].图书馆理论与实践,2011(6):72-76.
- [35] 张艳敏,马秀峰.中外数字资源长期保存策略比较研究[J].图书馆学研究,2009(6):29-32.

- [36] INGEBORG V. Networking for Digital Preservation: Current Practice in 15 National Libraries [M]. Munchen: K.G. Saur, 2006.
- [37] Public Consultation Document. Research Councils UK. Integrity, Clarity & Good Management - Code of Conduct and Policy on the Governance of Good Research Conduct [EB/OL]. [2012-06-08]. <http://www.rcuk.ac.uk/documents/reviews/grc/consultation.pdf>.
- [38] 董光彩.数字资源保存研究的发展历程及趋势[J].数字图书馆论坛,2006,31(12):36-39.

作者简介

胡译文 (1985-), 男, 南京大学信息管理学博士研究生, 研究方向: 科技信息的深层次挖掘、计量、评价与可视化分析。

武夷山 (1958-), 男, 研究员, 副所长, 研究方向: 科学计量学与科技评价。

Research Developments, Hotspots and Fronts for Digital Resources Preservation

Hu Zewen / Department of Information Management, Nanjing University, Nanjing, 210093

Wu Yishan / Institute of Scientific & Technical Information of China, Beijing, 100038

Sun Jianjun / Department of Information Management, Nanjing University, Nanjing, 210093

Abstract: Firstly, we discussed the concepts of digital resources preservation, and introduced its developmental history and current status of research at home and abroad. Then a visualizing analysis tool named "CitespaceII" and some visualizing analysis methods were applied to construct time-series evolution mapping and time-zone evolution mapping for domain knowledge basis (references), and drew domain's development trend chart, with them analyzing research developments of digital resources preservation. Besides, a mixed co-occurrence network between terms and keywords, and some co-cited cluster mappings for quotations were also constructed to identify research hotspots in digital resources preservation, as well as a detection method for burst terms was utilized to detect research fronts in digital resources preservation. Finally, research developments for long-term preservation of digital resources as a research hotspot often focused by scholars in Library, information science and archives administration discipline was reviewed from several important aspects: research topics, best practices, preservation period, preservation technology and methods, as well as preservation strategies.

Keywords: Digital resources preservation, Research developments, Research hotspots, Research fronts, Knowledge mapping, Long-term preservation for digital resources

(收稿日期: 2012-07-31)